

软件工程小组

立项建议书

项目名称：Honeycomb

项目组：

学号	姓名	手机	电子邮箱
516030910122	朱锦昊	13262607680	clivehaha@outlook.com
515260910022	曹金坤	13162096831	caojinkun@sjtu.edu.cn
516030910111	马轲	18217253971	13735887700@163.com
516030910141	谢添翼	13262905585	330281987@qq.com

2018 年 5 月

一、项目的合理性和市场前景

我们小组提出的“HoneyComb”项目，是一个基于 web 的对学术文献提供检索、交流、数据可视化和定制化等服务的项目。提出这个项目的动机是基于对当前学术文献查询、管理和交流领域中缺乏强大工具这一现状以及大量科研工作者日常需求难以被高效地满足这一不平衡的现状。

在现实的科研和学习中，在进行文献索引时，对于从业者（特别是刚刚进入某个领域或者细分课题的新人）如何“找到高质量的文档”以及如何“了解文献和自己的兴趣的课题的关联度”往往是一个比较棘手的问题，这是因为对于现在常见的会议 / 期刊论文，单纯阅读 abstract 对于老手们可能已经足够来把握论文的基本重点，但是对于新手们，不仔细阅读文章内的细节常常不能使其了解文章的贡献所在，而且即使细读了一篇文献，准确了解 related work 中的前人工作的性质和价值也会是个需要花费不小时间成本的工作。

除此之外，对于包括计算机科学在内的需要工程实践的学术领域，复现论文中的结果，是一个更棘手的问题，此时，即使论文结果是在统一的背景设置或者标准数据集的前提下被提出的，但是包括参数设置等等细节的问题都可能导致结果的难以复现。在当前情况下，人们只能依赖自己小圈子内的交流、反复的试错或者等待论文作者 / 第三方放出代码等复现论文的细节才能解决问题。因此，如果可以有一个平台让研究者们可以对特定的文献进行集中的讨论交流，会对他们节约时间有巨大的帮助。

最后，对于需要进行大量的文献阅读和研究的工作者来说，整理和记录自己的文献积累，构造他们之间的知识图谱是一个时间成本巨大但是完成后会有巨大收益的工作，而且这样的功能和社交属性的记录会极大地方便后来者的学习以及

同侪之间的交流。试想，对于刚刚了解一个领域，读了几篇入门论文的新手们来说，如果有一个前辈可以根据自己的经验构建出了一个关联的文献阅读路径和评论，那前者会受到多大的助益呀。

综上所述，对于我们的“HoneyComb”项目的主要功能：文献检索、学术知识图谱、学术交流和文献管理等功能，尽管当下的市场上已经对其中个别的细分领域有了优秀的产品，但是他们都没有对相关功能进行整合，并且缺乏专一的功能细化。综合我们自己的学习和科研实践，我们相信，HoneyComb项目不仅迎合了当下许多科研工作者的刚需，可以为他们带来巨大的收益，而且在当下的市场上缺乏足够有力的替代品和竞争者，市场前景广阔。

二、项目目标和特性

1. 项目目标

Honeycomb的目标客户群体是从事科研工作的学者以及在学习过程中需要阅读文献的学生。Honeycomb致力于给这一群体提供一个专业化和一站式的学术文献检索、管理、交流的平台。

在当前的市场上，对于文献的管理已经有Mendeley等软件做了非常优秀的工作，但是此类软件在学术交流方面缺少比较的社交属性；在我们比较熟悉的计算机科学领域，Github和WordPress等网站常被用作论文复现和论文交流的平台，但是这样的非专门化的平台难以满足更大范围和更高效的交流需要；对于知识图谱的建立，个人可以通过Xmind类的软件或者笔记应用来进行，但是这样的工作显然难以与别人进行分享。而对于我们的Honeycomb项目，文献网络图的设计既包含了文献搜索也拥有了知识图谱的全面性。同时，类似于Github，

CSDN 的社交设计便于科研从业者分享自己的学习经验 ,也有助于初学者从他人的文献网络图谱中更快的入门。因此 ,为 Honeycomb 在整合了现有的文献管理、学术分享网站功能的基础上 ,将论文文献与知识图谱进行结合 ,能够给科研从业者带来极大的便利。

2. 项目特性

2.1 功能性部分 :

1. 用户登录 (高)
2. 用户注册 (高)
3. 搜索框搜索论文 (高)
4. 网络图搜索论文 (高)
5. 浏览论文 (高)
6. 文献评论 (中)
7. 论文或者私人的论文仓库/知识图谱的收藏 (低)
8. 管理个人账号 (高)
9. 管理个人空间 (中)
10. 管理个人信息 (中)
11. 管理个人收藏 (低)
12. 管理个人网络图 (高)
13. 管理个人历史纪录 (中)

2.2 非功能性部分 :

1. 易用性 : 用户操作难度、用户界面、导航、图形标准 (高)
2. 可靠性 : 可用时间、平均故障间隔时间、精确率、最高缺陷率、错误 (高)

3. 性能：响应时间、吞吐量、容量、降级模式、资源利用情况（中）

4. 可支持性：编码标准、命名约定访问权限（中）

三、项目技术方案



整个系统由三个部分构成：

1. 数据管理模块：

数据模块的开发包括用户数据、论文数据、浏览数据三部分组成。其中用户数据和浏览数据保存在 MySQL 数据库中。而论文数据由于其庞大的数据量初步采用即时数据爬取和部分数据库存储。后端数据管理则采用 SSH 框架。

数据管理模块会对应一个爬虫子模块和一些相关第三方数据库的数据接口，这一部

分的实现会独立于网站的常规后台业务逻辑。

2. 分析处理模块：

在分析处理模块中，检索列表负责实现论文搜索，定制化负责管理用户个人信息以及私人网络图，关系分析则通过分析论文之间的关系数据构建网络图。

3. 展示体系模块：展示体系模块由 Web 门户、移动端 App、开发者 API 三部分组成。

Web 前端采用 Vue.js 框架，移动端 App 则采用微信小程序接口，而开发者 API 视后期开源需求而定。

四、项目风险分析和里程碑计划

项目风险：

1. 需求风险：在项目开发过程中，小组成员可能会对需求产生误解，另外随着软件开发的进行，项目组对需求的了解会更加深入，会导致需求的变更。而需求误解带来的返工和需求变更会对项目的进度、人力资源、质量产生很大的影响。因此，在项目启动阶段，小组成员之间应该充分交流以确保对需求的充分理解。与此同时，开展充分的需求调研，与项目目标客户群体进行定期的沟通。以此来保证软件需求在后续的开发过程中尽可能地稳定。
2. 进度风险：项目开发时间只有不到两个月时间，再加上小组成员开发经验缺乏，整个开发时间较为紧迫。因此采用三次冲刺（Sprint1、Sprint2、Sprint3）来增量式实现功能，以确保项目进度的合理性。
3. 技术和架构风险：在开发过程中，需要使用一些新的技术和新的架构。倘若架构设计不妥或者技术架构不熟悉会导致大规模的返工，如果过高估计新技术或方法带来的效率，会导致过于乐观的计划。因此在冲刺的第一阶段首要进行架构

的设计，并开发架构原型。

里程碑计划（参考 SCRUM 及进行设计）：

迭代	里程碑描述	成果
项目启动 7月1日~7月8日	需求调研与分析，开发可行性验证； 开发计划细化、确定开发的技术路径、技术框架和工具，开发界面原型； 项目立项	完成产品待办事项，即确定软件需求
Sprint 1 (两周) 7月9日~7月22日	架构分析与设计； 架构实现与搭建； R1的需求分析、设计与实现； 系统测试，并获得用户反馈，进行缺陷修复与改进。	完成系统版本 1(R1)的开发
Sprint 2 (两周) 7月23日~8月5日	在 R1 的基础上进行 R2 的需求分析、设计与实现； 系统测试，并获得用户反馈，进行缺陷修复与改进。	完成系统版本 2 (R2)的开发
Sprint 3(两周) 8月6日~8月19日	在 R2 的基础上进行 R3 的需求分析、设计与实现； 系统测试，并获得用户反馈，进行缺陷修复与改进。	完成系统版本 3 (R3)的开发
Sprint 4 (两周) 8月20日~8月31日	安装 R3，撰写用户手册，对用户进行培训，系统上线运行，并根据系统运行反馈进行缺陷修复与改进。	完成系统版本 3 的发布

功能实现：

1. 系统版本 1 (R1) 实现的功能：
 - a) 系统前端界面的主要模块
 - b) 第三方数据源的论文数据爬取
 - c) 后端数据库设计与构建

- d) 数据获取接口和数据存储模块的交互
- e) 用户注册与登录
- 2. 系统版本 2 (R2) 实现的功能 :
 - a) 论文网络图构建
 - b) 搜索框搜索论文实现
 - c) 论文浏览实现
- 3. 系统版本 3 (R3) 实现的功能 :
 - a) 用户个人空间创建与管理
 - b) 私人论文网络图谱构建
 - c) 线上用户创建的知识图谱和论文仓库的分享、收藏和评价
 - d) 站内注册用户的私人留言功能

五、 项目预期成果

- ✓ 项目计划
- ✓ 迭代计划
- ✓ 迭代评估报告
- ✓ SRS 文档和用例模型 (.oom)
- ✓ 软件架构文档和分析设计模型 (.oom)
- ✓ 测试用例和测试报告
- ✓ 工程代码
- ✓ 已上线的项目网站
- ✓ 微信小程序的项目工程

- ✓ 演示 demo 视频
- ✓ 演示 ppt